ベイズ推定による情報仮説の評価：その理論と各種モデルへの応用について

岡田謙介[1]

Recent advances in Bayesian evaluation of informative hypotheses: Its theory and applications to psychometric models

Kensuke Okada[1]

Abstract：Null hypothesis significance testing has been one of the most frequently used statistical techniques in psychological research. However, the hypotheses considered in null hypothesis significant testing are usually either unrealistic or uninformative. In contrast, in the context of Bayesian statistics, researchers can directly evaluate the researchers' informative hypotheses by simple arithmetical calculations. In this paper, first the basic theory and practice of Bayesian evaluation of informative hypotheses are presented. Next, the specification of prior distributions is discussed. The paper concludes with a review of applications, which includes multiple regression, contingency tables, structural equation modeling, and meta-analysis.
Keywords：Bayesian evaluation of informative hypothesis, Bayes factor, prior distribution

1.　はじめに

心理学をはじめとするさまざまな分野において，伝統的にデータ分析の手法としてはNeyman and Pearson（1933）流の帰無仮説検定（null hypothesis significance testing，以下仮説検定と呼ぶ）の枠組みが利用されてきた。これはデータから検定統計量の実現値とｐ値を求め，ｐ値が有意水準（典型的には0.05などの値に設定される）よりも小さければ帰無仮説を棄却し，本来研究者が主張したい対立仮説を採択する，という考え方に基づく方法論である。仮説検定は長きにわたりさまざまな分野のデータ分析を支えてきた有用な方法論のひとつであるが，しかし，それが本来持つ意味を越えた形で応用場面で使われてしまっているという批判もまた近年強まっている（Anderson, Burnham, & Thompson, 2000; Cohen, 1994; Cumming, 2012; Kline, 2013; 大久保・岡田，2012）。

心理学において，仮説検定に過度に依存することを避け，効果量や信頼区間など代替する指標を用いたデータ分析を推進する運動は，心理学における統計改革（statistical reform in psychology）と呼ばれることもある（Cumming et al., 2007; Finch et al., 2004）。心理学における統計改革の流れの中で，特に近年注目されている枠組みのひとつに，ベイズ推定による情報仮説の評価（Bayesian evaluation of informative hypotheses）がある（Gu, Mulder, Deković, & Hoijtink, 2014; Hoijtink, 2011, 2013）。ベイズ推定による情報仮説の評価は，研究者の持つ仮説を，母数に不等式制約を入れた情報仮説として表現し，直接評価するための枠組みである。次節で述べるとおり仮説検定では研究者の仮説についての確率が評価できないため，情報仮説の評価は仮説検定を補完・代替できる方法論として注目されている。

本稿では，ベイズ推定における情報仮説の評価について，その理論と考え方を説明し，特に心理統計学で用いられる４種類のモデルにおける利用について，最近の研究を具体的な数値例を含めてレビューを行う。

第２節では，情報仮説とは何かについて論じる。第３節では，情報仮説を評価する際に用いられる，ベイズファクターについて論じる。第４節では，ベイズ推定において利用することになる事前分布の設定について論じる。第５節では，重回帰分析・クロス集計表・構造方程式モデリング・メタ分析という心理統計学においてよく利用される四つのモデルをとりあげ，それぞれにおいてベイズ推定による情報仮説の評価を提案・利用する先行研究を，具体的な数値例とともにレビューする。第６節ではまとめと議論を行う。

2.　ベイズ推定による情報仮説の評価

いま例として，μ_1，μ_2，μ_3という３群の平均が関心の対象である場合を考える。心理学をはじめとする多くの

受稿日2014年11月25日　受理日2014年12月11日
1　専修大学人間科学部心理学科（Department of Psychology, Senshu University）
本研究は文部科学省私立大学戦略的基盤形成支援事業 S1101013，JSPS 科研費24730544，26285153の助成を受けたものです。

データ分析において利用される，古典的な仮説検定で検討する帰無仮説（null hypothesis）は，典型的に

$$H_n：\mu_1 = \mu_2 = \mu_3 \tag{1}$$

もしくは

$$H_n：\mu_1 = 0,\ \mu_2 = 0,\ \mu_3 = 0 \tag{2}$$

という形をしている。これらはいずれも，帰無仮説のもとでの母数の値を母数空間上の１点に定める単純帰無仮説（simple null hypothesis）である。これらに対して検定される対立仮説（alternative hypothesis）は

$$H_a：\text{not } H_n \tag{3}$$

つまり帰無仮説が真ではないというものである。帰無仮説と対立仮説は伝統的にはそれぞれ H_0, H_1 の記号で表現されることが多いが，本稿では情報仮説を H_1, H_2,… と表記するため，古典的な仮説検定における帰無仮説と対立仮説をそれぞれ H_n と H_a で表現することとする。

こうした単純帰無仮説を検定する仮説検定の枠組みについて，前節で述べたとおり統計改革の流れのなかでさまざまな批判が行われている。本稿の議論上特に重要なのは，古典的な仮説検定において評価しているのが，仮説の確率ではなく，仮説を所与とした（仮説で条件づけた）ときのデータの確率だということである。例えば，古典的な仮説検定で考える p 値とは，帰無仮説 H_0 が真である確率ではなく，帰無仮説 H_0 が真であることを仮定したときに得られるデータについての確率である。p 値はしばしば帰無仮説 H_0 が正しい確率であると誤解されているが，それは大きな誤りである。p 値をめぐるさまざまな誤解については Gelman（2013）に詳しい。

このような誤った解釈や理解が多いことは，逆説的にではあるがに p 値という量が，私たちが通常仮説の評価に期待する性質を満たしていないからだと考えられる。私たちは仮説で条件付けた確率ではなく，仮説の確率そのものを知りたいのである。

しかしながら，古典的な仮説検定においては仮説が正しい確率は知ることができない。なぜならば，帰無仮説検定は立脚しており，そして頻度論においてはデータが確率変数・母数が定数であるからである。統計学で検証の対象となる H とは母数についての仮説であり，頻度論ではその枠組み上，仮説が正しい確率を考えることができないのである。

それに対して，統計学において頻度論と並び立つもう一つの枠組みであるベイズ統計学では，逆にデータが確率変数で母数が定数であると考える。このとき，母数の確率や，母数の関数である仮説の確率 $p(H|X)$ を考えることが可能になる。これを利用して研究者の持っている情報仮説を，ベイズ統計の枠組みで評価しようというのがベイズ推定による情報仮説の評価の動機である。

情報仮説（informative hypothesis）は，分析者の仮説を不等式制約によって直接表現する仮説である。いま，第１群よりも第２群の平均が大きく，それよりも第３群の平均の方が大きいという仮説を研究者が持っているとしよう。これは，母数についての不等式制約として

$$H_1：\mu_1 < \mu_2 < \mu_3 \tag{4}$$

と表現することができる。また，母数の一部については制約をいれない情報仮説も可能である。例えば，第１群の母平均よりも第２群や第３群の平均が大きいことは言えるが，第２群と第３群の間については特にわからないという仮説を持っているとする。これは，

$$H_2：\begin{cases}\mu_1 < \mu_2 \\ \mu_1 < \mu_3\end{cases} \tag{5}$$

いう仮説として表現することができる。こうした情報仮説のよさを，直接データに基づいて評価しようというのがベイズ推定による情報仮説の評価の考え方である。

これらの情報仮説は，パラメータ空間の部分空間を指定していることに注意する。母数のとる値に関して制約をいれない仮説を無制約仮説（unconstrained hypothesis）

$$H_0：\mu_1,\ \mu_2,\ \mu_3 \tag{6}$$

とする。この無制約仮説 H_0 と整合的な空間を図示すると，Figure １のとおり母数空間全体である。ただし，ここで目盛は各母数のとりうる値についての最小値から最大値までの割合を表すものとする（本説の以下の図でも同様）。

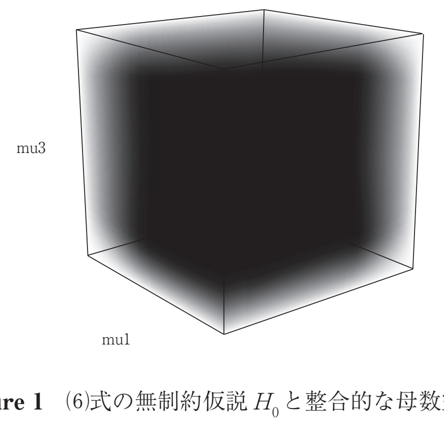

Figure 1　⑹式の無制約仮説 H_0 と整合的な母数空間

このうち，情報仮説 H_1 と整合的な母数空間はそれぞれ Figure 2, Figure 3 に示すとおりの部分空間である。情報仮説 H_1 と整合的な母数空間は，無制約仮説 H_0 と整合的な母数空間のうち１/６であり，また情報仮説 H_2 に

ついては同じく 1／3 であることがわかる。このように，情報仮説と整合的な母数空間が，無制約仮説の母数空間の部分空間となることは重要である。

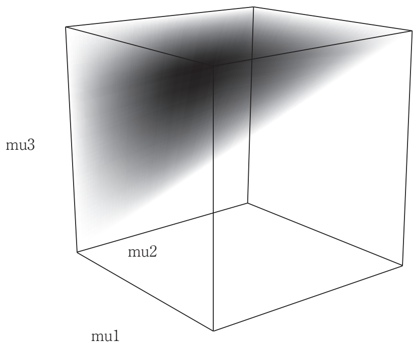

Figure 2 ⑷式の情報仮説 H_1 と整合的な母数空間

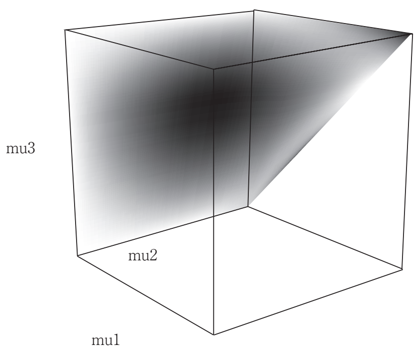

Figure 3 ⑸式の情報仮説 H_2 と整合的な母数空間

なお，この観点から⑴式もしくは⑵式の帰無仮説 H_a と整合的な母数空間を図示してみると Figure 4 のようになる。Figure 4 には（事実上見えないかも知れないが）⑴式もしくは⑵式の条件を満たす 1 点がプロットされている。このように古典的な帰無仮説検定における帰無仮説とは母数空間のうち唯一の 1 点であり，それと整合的

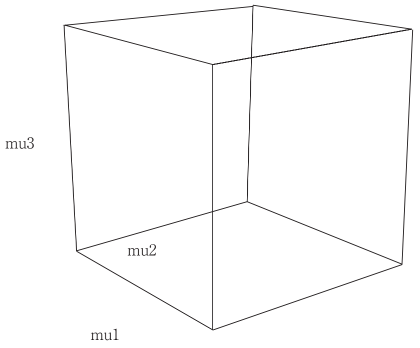

Figure 4 古典的な帰無仮説 H_n と整合的な母数空間

な値を母数が正確にとることは事実上考えられない仮説である。このことから，Cohen（1994）は帰無仮説を実用上正しいことはありえない「皆無仮説（nil hypothesis)」だとしている。

一方で，対立仮説 H_a と整合的な母数空間とは，無制約仮説の母数空間のうち H_n で指定した 1 点をのぞくすべての部分であり，事実上 Figure 1 に示した無制約仮説と整合的な母数空間と等しい。すなわち，古典的な仮説検定において採択しようとする対立仮説は，実はほとんど情報を持たない仮説である。

これらのことから，Hoijtink, Klugkist, and Boelen（2008）は，仮説検定における帰無仮説は現実的でない（unrealistic）仮説であり，一方の対立仮説は情報のない（uninformative）仮説であると批判している。

そこで，ベイズ推定による情報仮説の評価の枠組みでは，帰無仮説を棄却して対立仮説を採択する代わりに，無制約仮説と比較したときの情報仮説のよさをベイズファクターによって評価する。この枠組みについてより詳しくは，Hoijtink（2011）および Hoijtink et al.（2008）を参照いただきたい。

3. ベイズファクター

ベイズファクター（Bayes factor）は Jeffreys（1935）の提案による，ベイズ統計学での仮説の評価のための指標である。ベイズファクターはしばしば，ベイズ比，ベイズ因子などと呼ばれることもある。いまデータ X に基づいて仮説 H_i を仮説 H_j と比較することを考えるとき，ベイズファクターは以下のように定義される。

$$BF_{ij} = \left(\frac{p(H_i|X)}{p(H_j|X)}\right) \Big/ \left(\frac{p(H_i)}{p(H_j)}\right) \tag{7}$$

⑺式の分子はデータを得た後の事後確率の比であり，分母は同じく事前確率の比である。確率の比のことをオッズ（odds）と呼ぶ。従って，ベイズファクターは，仮説 H_i と H_j のオッズが，データ X を得る前から得た後でどう変化したかを比によって表現した量である。BF_{ij} が 1 より大きければ，データによって仮説 H_i のほうがより支持されたということになり，逆に 1 より小さければ仮説 H_j の方がより支持されたということになる。

古典的な仮説検定における p 値について，有意水準 5％が慣習的なカットオフ基準となっているのと同じように，ベイズファクターについても慣習的な基準が存在する。よく知られているものに，Jeffreys（1961）の基準と Kass and Raftery（1995）の基準がある。それぞれ

を Table 1に示す。なおここでの日本語訳は大森・渡部（2008）を参考にした。

Table 1 ベイズファクターについてよく知られた2つの基準

(a) Jeffreys（1961）

BF_{ij}	H_j に反する証拠の大きさ
1 to 3.2	あまりあるとは言えない
	(Not worth more than a bare mention)
3.2 to 10	十分ある（Substantial）
10 to 100	強くある（Strong）
>100	決定的にある Decisive

(b) Kass and Raftery（1995）

BF_{ij}	H_j に反する証拠の大きさ
1 to 3	あまりあるとは言えない
	(Not worth more than a bare mention)
3 to 20	肯定的である（Positive）
2 to 15	強くある（Strong）
>15	非常に強くある（Very strong）

ベイズ推定における情報仮説の評価では，(6)式のような無制約仮説 H_0 に対して(4)式や(5)式のような情報仮説 H_i のよさを比較する。このときのベイズファクターは

$$BF_{i0} = \frac{f_i}{c_i} \tag{8}$$

という非常にシンプルな式によって与えることができる。ここで，c_i と f_i は，それぞれデータを得る前（事前）と得た後（事後）における，無制約仮説 H_0 と整合的母数空間に占める情報仮説 H_i と整合的な母数空間の割合として定義される。c_i は情報仮説 H_i の相対複雑さ（relative complexity）と呼ばれ，無制約仮説 H_0 に比べてどれだけ制約の強い仮説であるのか，別の言葉でいうとどれだけ情報の多い仮説であるのかを表す。小さいほど制約が強く，また情報が多いことに注意する。f_i は情報仮説 H_i の相対当てはまり（relative fit）と呼ばれ，データに基づき情報仮説がどれだけ支持されたのかを表す（Mulder, 2014a）。

ベイズ推定による情報仮説の評価では情報仮説を無制約仮説と比較する(8)式の BF_{i0} が利用されることが多かったが，近年 Mulder, Hoijtink, and Klugkist（2010）によってが BF_{i0} 上に有界であることが指摘され，代わりに情報仮説とその補仮説（composite hypothesis）

$$H_{i_c}:\ \text{not}\ H_i \tag{9}$$

との比較が用いられることも多くなっている。情報仮説 H_i とその補仮説 H_{i_c} を比較するベイズファクター BF_{ii_c} は，

$$BF_{ii_c} = \frac{f_i}{c_i} \Big/ \frac{(1-f_i)}{(1-c_i)} \tag{10}$$

とやはりシンプルな形で与えることができる。

ところで，ベイズ推定において第一義的に関心があるのは通常母数の事後確率であるが，情報仮説の評価においてはなぜ事後確率ではなくベイズファクターがよく利用されるのだろうか。それは，仮説 H_i の事後確率 $p(H_i|X)$ の値は，ほかに何種類の仮説を考えるかによって変わるからである。いま，考えられる仮説が $H_0, H_1, \ldots, H_m$ の m とおりであるとき，情報仮説 H_i の事後確率は

$$p(H_i|X) = \frac{BF_{i0}}{1+\sum_{j=1}^{m} BF_{j0}} \tag{11}$$

と与えることができる。なお無制約仮説 H_0 の事後確率は

$$p(H_0|X) = \frac{1}{1+\sum_{j=1}^{m} BF_{j0}} \tag{12}$$

となる。

(11)式からわかるように，情報仮説 H_i の事後確率は，一般に代替案となる仮説の個数 m が大きいほど小さくなる傾向がある。しかし，ベイズファクターは2つの事後確率だけを比較するため，ほかにいくつ可能な仮説があるかに依存しない。そのため，比較対象として無制約仮説 H_0 もしくは補仮説 H_{i_c} を採用した，ベイズファクター BF_{i0}, BF_{ii_c} が利用しやすいのである。もちろん，考える仮説の数 m が所与のもとでは事後確率 $p(H_i|X)$ も有益な指標である。

4. 事前分布の設定について

ベイズ統計学においては，母数を確率変数と考え，データを取得しベイズの定理を用いて母数の事前分布を事後分布へとアップデートする。この枠組みにおいて，まず母数に事前分布を設定する必要がある。

もちろん，事前情報がある場合には，その情報を事前分布に利用すればよい。しかし，必ずしも事前情報がないことも多い。このような場合において事前分布をどのように設定すればよいかが問題となる。

Hoijtink（2013）は，考えている情報仮説と整合的な母数空間が，無制約仮説の母数空間を等しい D 組の部分空間に分割したうちのひとつである場合，すなわち

D 個の同等集合（equivalent set）のうちの一つである場合には，ベイズファクターによる情報仮説の評価に，以下のよい性質があることを示した。

(1) 事前分布に適切な無情報事前分布を設定することにより，仮説の相対当てはまり f_i が本質的にデータの情報から決まること

(2) 従って，このとき，事前知識を入れ込まない「客観的な」情報仮説のベイズファクターを求められること

ここでの前提条件について，より厳密には Hoijtink (2013) を参照してほしい。

上記の二つの性質は研究者の持っている仮説を評価する上で望ましいと考えられるものであるが，すべての情報仮説がこの前提条件を満たすわけではないことにも注意が必要である（反例も Hoijtink の同論文中に紹介されている）。本稿で扱う情報仮説はすべてこの前提条件を満たすものである。

ベイズ推定による情報仮説の評価における事前分布の設定については，客観ベイズの枠組みなどを利用した理論研究が近年進んでおり，よりさまざまな自体で汎用的に利用できる手法の開発が期待される（Mulder, 2014a, 2014b)。

5. 各種モデルへの応用例

情報仮説の評価はさまざまな枠組みに応用することができる。本節では，心理学でよく利用されるモデルを中心に四つの統計モデルについて，最近の情報仮説の評価の研究をレビューする。岡田（2014）が分散分析における情報仮説の評価をとりあげているため，本稿ではそれ以外のモデルをとりあげる。

5.1 重回帰分析・線形モデル

回帰分析を含む線形モデルにおけるベイズ推定における情報仮説の評価は Mulder et al. (2010) により提案された。ここでは Kluytmans, van de Schoot, Mulder, and Hoijtink (2012) による分析例を紹介する。彼らは提案した方法論の応用場面として，過食（overconsumption）を従属変数 y，情動的摂食（emotional eating）と抑制的摂食（restrained eating）という食行動の 2 変数を説明変数 x_{emo}, x_{res} とした重回帰分析の場面を考えた。重回帰モデルは

$$y = \beta_0 + \beta_{emo}x_{emo} + \beta_{res}x_{res} + \varepsilon \tag{13}$$

として表現される。

ここで彼らの考えた情報仮説は 3 種類ある。まずはじめに，情動的摂食と抑制的摂食のそれぞれ独自効果が過食に制の影響を与えるという情報仮説

$$H_1: \beta_{emo} > 0,\ \beta_{res} > 0 \tag{14}$$

である。第二に，情動的摂食の独自効果は抑制的摂食の独自効果よりも大きいという情報仮説

$$H_2: \beta_{emo} > \beta_{res} \tag{15}$$

である。第三に，この H_1 と H_2 がともに成り立つという情報仮説

$$H_3: H_1 \cap H_2 \tag{16}$$

である。なお H_3 は母数に不等式制約をいれた形では，次のようにも表現することができることがわかる。

$$H_3: \beta_{emo} > \beta_{res} > 0 \tag{17}$$

Kluytmans et al. (2012) はこれら三つの情報仮説を，Van Strien, Herman, and Verheijden (2009) によるオランダでの過食と摂食行動に関するデータセットを使って評価した。同論文中で報告されたベイズファクター BF_{i0} と，そこから求めた事後確率 $p(H_i|X)$ およびベイズファクター BF_{ii_c} を Table 2 に示す。ベイズファクターはほぼ 1 であり，各モデルの事後確率は小数点以下 2 位に丸めるといずれも0.25つまりと 1/4 なっている。従って，このデータからは，四つの仮説いずれをも強く支持する結果は得られず，また四つの仮説それぞれの確からしさと無情報仮説の確からしさに，総じてほぼ全く差がないことがわかる。

Table 2 van Strein et al. (2009) の過食データにおける重回帰分析での情報仮説の評価（Kluytmas et al., 2012 の結果を利用）[1]

	H_0	H_1	H_2	H_3
c_i	–	0.247	0.49	0.12
f_i	–	0.25	0.50	0.12
$p(H_i\|X)$	0.25	0.25	0.25	0.25
BF_{i0}	1	1.01	1.01	0.99
BF_{ii_c}	1	1.02	1.04	1.00

なお，Kluytmans et al. (2012) ではこのほかに，仮に H_1 や H_3 が真であることを前提として模擬乱数データを発生させた場合のベイズファクターの値も報告されており，その場合にはベイズファクターは真の仮説を適切に支持することが示されている。

5.2 分割表の分析

Laudy and Hoijtink (2007) は分割表における情報仮説の評価を提案した。彼らがとりあげた応用例は，1 日

のアルコール消費量と１日当たりの喫煙量による口腔ガンの発生率についての，２×２のクロス表である（Table 3)。ここで π_{ij} は，各クロス表におけるケース群とコントロール群との発生率の比を表すとする。π_{ij} が１より大きいときはケース群でより発生率が高く，１より小さいときはその逆であることを意味する。

Table 3　Laudy and Hoijtink（2007）における情報仮説のための２×２クロス表

喫煙量（本／日）	アルコール摂取量（オンス／日） 0	0.1-3	0.4-1.5	1.6＋
0	π_{11}	π_{12}	π_{13}	π_{14}
1-19	π_{11}	π_{12}	π_{13}	π_{14}
20-39	π_{11}	π_{12}	π_{13}	π_{14}
40＋	π_{11}	π_{12}	π_{13}	π_{14}

彼らはこのクロス表にまとめられる口腔ガンの発生率データについて，以下の３種類の情報仮説を設定した。第１に，アルコール摂取量が多いほど，ガン発生率が高くなるという仮説である。第２に，喫煙量が多いほどガン発生率が高くなるという仮説である。第３に，これら二つががいずれも真であるという仮説である。それぞれは情報仮説として以下のように表すことができる。

$$H_1: \pi_{i1} < \pi_{i2} < \pi_{i3} < \pi_{i4} \quad (\text{for } i=1, \ldots, 4) \tag{18}$$

$$H_2: \pi_{1j} < \pi_{2j} < \pi_{3j} < \pi_{4j} \quad (\text{for } j=1, \ldots, 4) \tag{19}$$

$$H_3: H_1 \cap H_2 \tag{20}$$

これらの仮説に無制約仮説 H_0 を加えた四つの仮説が，ベイズ推定による情報仮説の評価を用いて比較検討された。データとしては Rothman and Keller（1972）によるケースコントロール研究のものが利用された。Laudy and Hoijtink（2007）で報告された各モデルの事後確率 $p(H_i|X)$，およびそこから算出したベイズファクター BF_{i0}，BF_{ii_c} を Table 4 に示す。最も制約の強い情報仮説 H_3 が最もデータから支持されており，その無制約仮説と比較したベイズファクターは26と情報仮説 H_3 を強く支持する証拠があることが見てとれる。

なお，実際には Laudy and Hoijtink（2007）の提案による方法は2, 3節で説明した，c_i と f_i を利用する現在一般的な情報仮説の評価の方法とは異なったアプローチをとっている。そこで，ここでは彼らが報告する事後確率と，そこから直接導かれるベイズファクター BF_{i0} のみを示した。より一般的な枠組みでの分割表の分析については Hoijtink（2011, Chapter 7）を参照するとよい。

Table 4　Laudy and Hoijtink（2007）におけるアルコールと喫煙が口腔ガンに与える影響の研究についての情報仮説の評価（Rothman and Keller, 1972のデータを利用）

	H_0	H_1	H_2	H_3
$p(H_i\|X)$	0.02	0.16	0.30	0.52
BF_{i0}	1	8	15	26

5.3　構造方程式モデリング

van de Schoot, Hoijtink, Hallquist, and Boelen（2012）は，構造方程式モデリング（structural equation modeling；別名 共分散構造分析）におけるベイズ推定における情報仮説の評価を提案した。構造方程式モデリングはモデルの柔軟性が高く研究者が自由に仮説をパス図の形で表現し分析できることがよく知られた利点であるが，しかし情報仮説の評価は典型的な構造方程式モデリングには含まれていなかった。彼らは，Boelen and van den Bout（2008）による，複雑性悲嘆と非複雑性悲嘆の弁別のためのデータを用いて，Figure 5の形の構造方程式モデルを検討した。ここで「複雑性悲嘆」と「非複雑性悲嘆」はそれぞれ複数の観測変数によって測定される潜在変数であるが，このパス図では測定変数や誤差は略し，構造方程式部分のみを示している。

彼らの検討した情報仮説は

$$H_1: \begin{cases} \lambda_1 - \lambda_3 > 0 \\ \lambda_2 - \lambda_4 > 0 \end{cases} \tag{21}$$

というもの，つまり複雑性悲嘆からの従属変数へのパス係数が，非複雑性悲嘆からのそれよりも，２従属変数いずれについても大きいという仮説である。

同論文中では，(21)式の情報仮説についての c_i，f_i，BF_{ii_c} が報告されている。さらに BF_{i0} と事後確率 $p(H_i|X)$ もここから導出し，併せて Table 5 に示した。情報仮説についてのベイズファクターの値は BF_{i0} が 2.796，BF_{ii_c} が6.967と確定的に大きいわけではないが，

Table 5　van de Schoot et al（2012）における，Bolen and van den Bout（2008）の構造方程式モデリングとデータを用いた情報仮説の評価

	H_0	H_1
c_i	−	0.25
f_i	−	0.699
$p(H_i\|X)$	0.263	0.737
BF_{i0}	1.000	2.796
BF_{ii_c}	1.000	6.967

しかし制約を入れない仮説と比べれば事後モデル確率の意味で4倍近くは支持されたことがわかる。

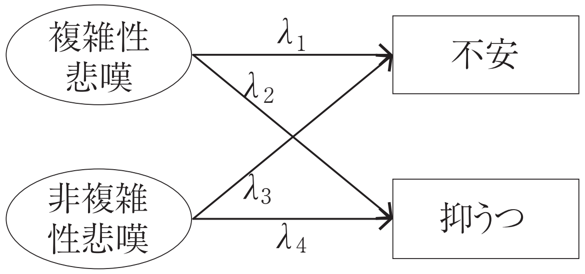

Figure 5 van de Schoot et al (2012) で利用された構造方程式モデリングのパス図（構造方程式モデル部分のみを示している）

5.4 メタ分析

Okada（*in press*）はメタ分析における情報仮説の評価を提案した。同論文では，テストの信頼性の指標として最もよく利用されるクロンバックの α について，α が平均的にある基準値よりも大きいという情報仮説を検討する場合，および複数のテストフォームについての平均的な α に関する情報仮説を検討する場合の二つの事例が含められている。ここでは後者をとりあげる。

マスラック・バーンアウト測定尺度（Maslach burn-out inventory, MBI）は，1970年代以降対人サービス職者の職業性ストレス反応として注目されるようになったバーンアウト現象について測定するための心理尺度である（荻野・増田 , 2004）。MBI ははじめ英語版が開発されたが，その後多くの言語に翻訳され，特にスペイン語版についてはその利用を報告する研究が多い。Aguayo, Vargas, de la Fuente, and Lozano（2011）がメタ分析のために MBI の情緒的疲弊感（emotional exhaustion）下位尺度について収集・整理した先行研究のデータを用いて，Okada（*in press*）は情報仮説を評価するメタ分析を行った。検討した情報仮説は，他国語版よりもスペイン語版の α の方が平均的に高く，さらにそれよりも英語版が平均的に高いという情報仮説

$$H_1 : \bar{\alpha}_{Eng} > \bar{\alpha}_{Spa} > \bar{\alpha}_{Other} \tag{22}$$

である。ベイズファクターおよび事後確率は Table 6 に示すとおりであった。ベイズファクターより，無制約仮説と比較した情報仮説の事後オッズは，データを得る前と比較して4.884倍になっていること，また補仮説と比較したときには同じく21.87倍になっていることがわかる。このことから，(22)式の情報仮説は先行研究の情報を用いたメタ分析によって支持されたと考えることができる。

Table 6 Okada（*in press*）による，Aguayo et al.（2011）の MBI のメタ分析での情報仮説の評価

	H_0	H_1
c_i	–	0.167
f_i	–	0.814
$p(H_i \mid X)$	0.170	0.830
BF_{i0}	1.000	4.884
BF_{ii_c}	1.000	21.87

6. まとめと議論

本稿では帰無仮説検定の欠点を克服し，補完・代替する方法として近年注目されているベイズ推定による情報仮説の評価について，その考え方を論じ，また同手法を用いて最近提案されている各種心理統計学のモデルにおける情報仮説の評価法について，具体的な数値例をとりあげつつ概観した。取り上げたモデルは，重回帰分析（線形モデル），分割表の分析，構造方程式モデリング，メタ分析であった。

ベイズ推定による情報仮説の評価は基本的にモデルに依存せず広く柔軟に利用することのできる方法論であり，特に潜在変数を用いたモデルなど心理統計学関連ではさらなる方法論の発展がのぞまれる。また，応用研究の進展も強く期待される。

なお，ベイズ統計学における，情報仮説の評価とは異なる仮説やモデルの評価の枠組みとして，予測を重視する考え方がある。この両者は立場を異にするものではあるが，どちらも実データ分析において相補的に利用できるものと考えられる。事後予測チェックや情報量規準 WAIC を用いた，予測の立場からのベイズモデルの評価についても，近年多くの研究が進んでいる。こうしたアプローチについては Gelman et al. (2013) を参照してほしい。

注

1 相対複雑さ c_i は今回のような情報仮説については理論的に求めうる値であるが，Kluytmans et al.（2012）では汎用ソフトウェアを用いて c_i はモンテカルロ法によって推定しているため理論的な値とは少しのずれがある。本稿では同論文中で報告された値をそのまま用いることにする。

7. 引用文献

Aguayo, R., Vargas, C., de la Fuente, E. I., & Lozano, L. M. (2011). A meta-analytic reliability generalization study of the Maslach Burnout Inventory. *International Journal of Clinical and Health Psychology*, **11**(2), 343-361.

Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management*, **64**(4), 912-923.

Boelen, P. A., & van den Bout, J. (2008). Complicated grief and uncomplicated grief are distinguishable constructs. *Psychiatry research,* **157**(1), 311-314.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist,* **49**(12), 997-1003.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., Kleinig, A., Lo, J., McMenamin, N., & Wilson, S. (2007). Statistical reform in psychology is anything changing? *Psychological Science*, **18**(3), 230-232.

Finch, S., Cumming, G., Williams, J., Palmer, L., Griffith, E., Alders, C., Anderson, J., & Goodman, O. (2004). Reform of statistical inference in psychology: The case of Memory & Cognition. *Behavior Research Methods, Instruments, & Computers*, **36**(2), 312-324.

Gelman, A. (2013). Commentary: P values and statistical practice. *Epidemiology*, **24**(1), 69-72.

Gelman, A., Carlin, B. P., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3 ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian Evaluation of Inequality Constrained Hypotheses. *Psychological Methods*, **19**(4), 511-527.

Hoijtink, H. (2011). *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. Boca Raton, FL: Chapman and Hall/CRC.

Hoijtink, H. (2013). Objective Bayes Factors for Inequality Constrained Hypotheses. *International Statistical Review*, **81**(2), 207-229.

Hoijtink, H., Klugkist, I., & Boelen, P. (2008). *Bayesian Evaluation of Informative Hypotheses*. New York: Springer.

Jeffreys, H. (1935). Some Tests of Significance, Treated by the Theory of Probability. *Proceedings of the Cambridge Philosophy Society*, **31**, 203-222.

Jeffreys, H. (1961). *Theory of probability* (3 ed.). Oxford, UK.: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**(430), 773-795.

Kline, R. B. (2013). *Beyond Significance Testing: Statistics Reform in the Behavioral Sciences* (2 ed.). Washington, DC: American Psychological Association.

Kluytmans, A., van de Schoot, R., Mulder, J., & Hoijtink, H. (2012). Illustrating Bayesian Evaluation of Informative Hypotheses for Regression Models. *Frontiers in Psychology*, **3**, PMC3265142.

Laudy, O., & Hoijtink, H. (2007). Bayesian methods for the analysis of inequality constrained contingency tables. *Statistical Methods in Medical Research*, **16**(2), 123-138.

Mulder, J. (2014a). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification. *British Journal of Mathematical and Statistical Psychology*, **67**(1), 153-171.

Mulder, J. (2014b). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. *Computational Statistics & Data Analysis*, **71**, 448-463.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, **140**(4), 887-906.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **231**, 289-337.

大久保街亜・岡田謙介 . (2012). 伝えるための心理統計：検定力・効果量・信頼区間 : 勁草書房 .

大森裕浩・渡部敏明. (2008). MCMC 法とその確率的ボラティリティ変動モデルへの応用. 小西貞則・国友直人（編）21世紀の統計学 I 社会・経済と統計科学. 東京大学出版会. pp. 223-266.

Okada, K. (*in press*). Bayesian meta-analysis of Cronbach's alpha to evaluate informative hypotheses. *Research Synthesis Methods*

岡田謙介 . (2014). ベイズ統計による情報仮説の評価は分散分析にとって代わるのか？ 基礎心理学研究 , **32**(2), 223-231.

荻野佳代子・増田真也 . (2004). 日本版 MBI-GS（Maslach Burnout Inventory-General Survey）の妥当性の検討 . 心理学研究 , **75**(5), 415-419.

Rothman, K., & Keller, A. (1972). The effect of joint exposure to alcohol and tobacco on risk of cancer of the mouth and pharynx. *Journal of Chronic Diseases*, **25**(12), 711-716.

van de Schoot, R., Hoijtink, H., Hallquist, M. N., & Boelen, P. A. (2012). Bayesian Evaluation of Inequality-Constrained Hypotheses in SEM Models Using M plus. *Structural Equation Modeling: a multidisciplinary journal*, **19**(4), 593-609.

Van Strien, T., Herman, C. P., & Verheijden, M. W. (2009). Eating style, overeating, and overweight in a representative Dutch sample. Does external eating play a role? *Appetite*, **52**(2), 380–387.